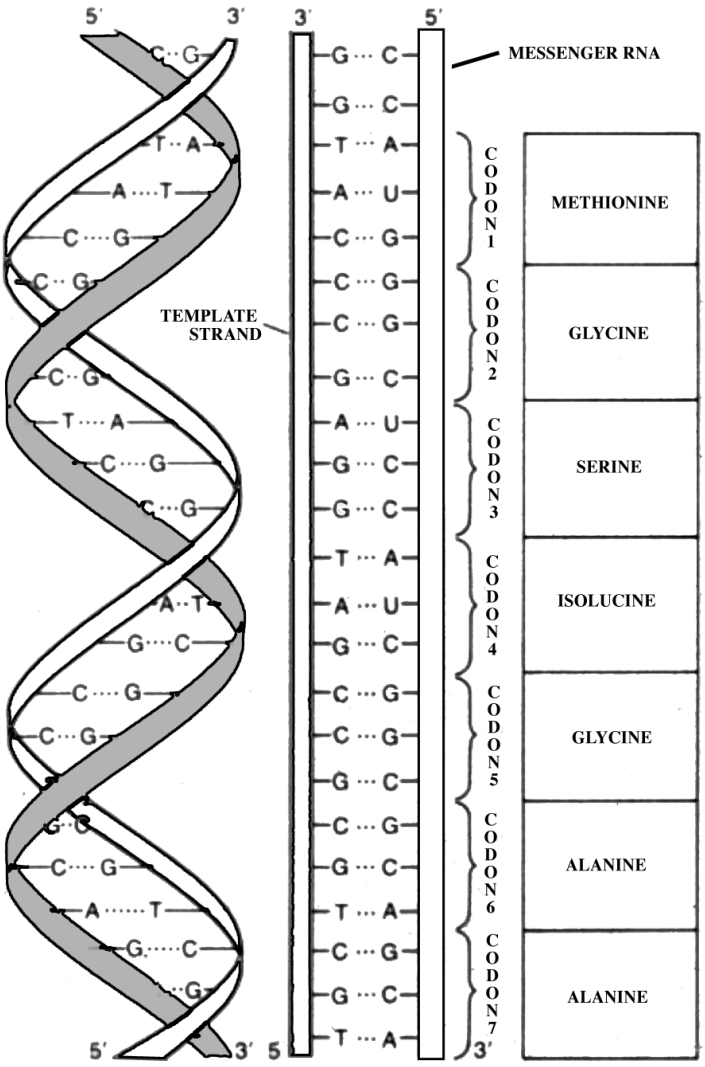
GENE EXPRESSION

Overview

The term gene expression often refers to the production of mRNA from the DNA template. However, in order for the expressed gene to actually result in a functional or phenotypic change, the mRNA must be translated into protein. The overall process is depicted to the left. The process of converting DNA to mRNA is called transcription. The process of converting mRNA into protein is called translation.



Transcription = the generation of single stranded RNA that is identical in sequence with the coding strand of DNA.

Transcription unit = a sequence of DNA that can be transcribed by RNA polymerase into a single RNA beginning at a startpoint (**initiation**) and end at a **terminator**.

Upstream sequences are 5' (prior) to the startpoint of transcription, **downstream** sequences are 3' (after) a particular sequence.

In the process of transcription, related genes often are transcribed simultaneously.

Gene family = a set of genes whose exons are related

> Gene families are believed to have arisen by duplication of an ancestral gene with different members of the family diverging as a consequence of mutations occurring during evolution. This may result in evolution of proteins that are optimized for specific functions (e.g., fetal hemoglobin has a higher affinity for oxygen than adult hemoglobin). Alternatively, mutations may lead to inability to produce a functional protein. These latter genes are referred to as **pseudogenes.**

Gene cluster = a group of genes that are adjacent and are identical or related to each other
Housekeeping genes = genes that encode proteins that provide basic functions in all cells and are always expressed (**constitutive**ly expressed); expression level is constant

RNA Polymerases

RNA polymerases are essential in the initiation of transcription. While there are three different RNA polymerases, RNA polymerase II is the type involved in gene expression that ultimately results in production of protein.

RNA polymerase I -- located in the nucleolus; transcribes genes that encode ribosomal RNA
RNA polymerase II -- located in the nucleoplasm; produces heterogeneous RNA that becomes mRNA after processing and splicing
RNA polymerase III -- located in the nucleoplasm; synthesizes small RNA such as tRNA

There are several steps to consider in understanding gene expression. These are tabulated below as control points for gene expression.

Potential control points for gene expression:
Activation of gene structure
Initiation of transcription
Elongation of transcript

Capping the RNA transcript
Splicing the RNA transcript
Cleaving the RNA transcript
Polyadenylation of the RNA transcript
Transport into the cytoplasm
Susceptibility of RNA to degradation
Translation of mRNA into protein
Post-translational modification of protein
Transport (and secretion) of the protein
Proteolytic cleavage of the protein
Interaction of protein with natural inhibitors

Gene activation = the state of DNA that allows it to be transcribed
• Structural changes in chromatin - unwinding of DNA strands
• Transcribable DNA has increased sensitivity to DNase
• Transcribable DNA is often undermethylated
CpG-rich islands consisting of mostly unmethylated CpG are often found upstream of consitiutively transcribed genes

Transcription

RNA polymerases are necessary but not sufficient for transcription. RNA polymerases bind to the **promoter** region of DNA upstream of the gene that will be transcribed. Accessory factors, called **transcription factors**, are also required.

Promoter = region of DNA where binding of RNA polymerase occurs to initiate transcription
Promoters are **cis-acting** sites on DNA

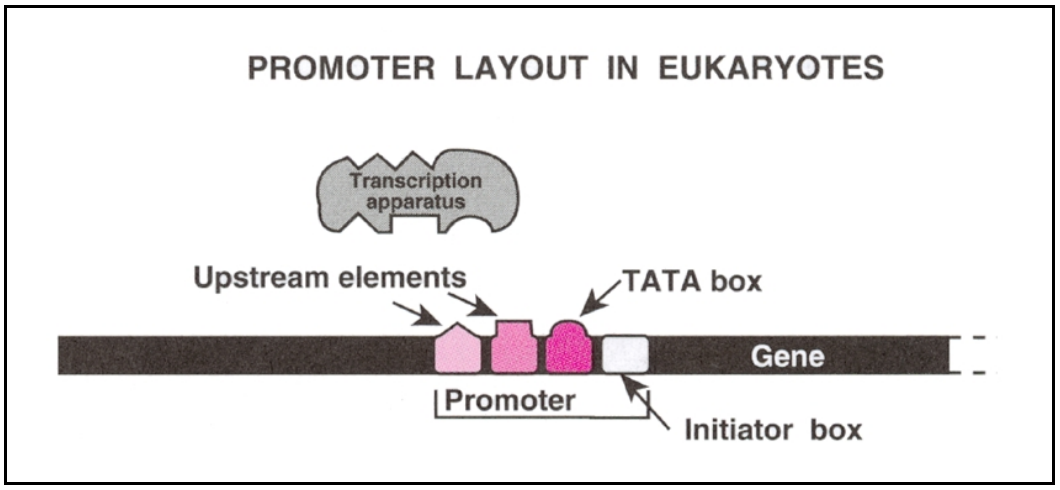


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, Illlinois, USA.

Promoters contain short conserved sequences recognized by RNA polymerase and different transcription factors

Hence, transcription factors are **tran-acting** factors that recognize certain cis-acting sites in the promoter

The complex of transcription factors that bind to specific promoter sequences in a particular gene are believed to control the initiation of transcription.

Promoters are generally greater than 100 base pairs, contain several conserved DNA sequences, are upstream but relatively close to the startpoint of initiation, and are protected from nuclease digestion by the binding of trans-acting factors.

Short **consensus sequences** within promoters are called **boxes** and function to help position polymerase correctly help determine the strength of the promoter.

TATA box (consensus sequence = TATAAAA)
CAAT box (consensus sequence = GGCCAATCT)
GC box (consensus sequence = GGGCGG)

Enhancers

First described as 72 base pair repeat elements in SV40
Enhance initiation

May be located relatively far from the startpoint
May be upstream or downstream from the startpoint
May be bound by transcription factors
May be in either orientation
May reside within introns
Consensus sequences are contiguous over short regions
Have a dense concentration of protein binding sites
Even when moved far away from a gene, enhancers can still enhance transcription

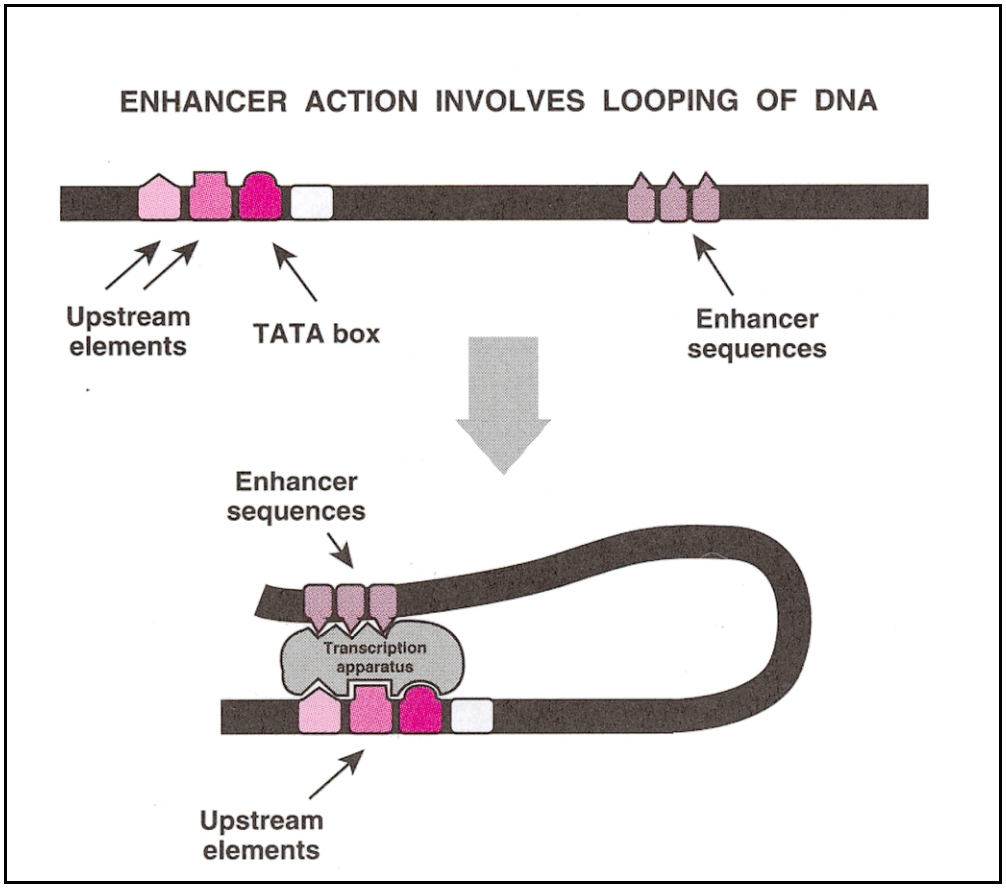


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Response Elements

The consensus sequences that uniquely identify groups of genes
These groups of genes are under common control
May share consensus sequences recognized by specific transcription factors
Transcription factors that recognize a response element coordinate the transcription of all the genes that have that response element

Are found within promoters and enhancers

Heat shock response element (HSE) - These response elements are localized in the promoter region of heat shock genes and are turned on in response to heat shock.

Glucocorticoid response element (GRE) - These response elements govern the response to steroid hormones and are typically located about 250 bp upstream from the startpoint of transcription.

Metal response element (MRE) enhance transcription of genes in response to heavy metals. An example of a gene with an MRE is the metallothionein gene.

TPA (phorbol ester) response element (TRE). TPA is a tumor promoting agent classically used in skin painting carcinogenicity studies. In response to TPA, AP1 transcription factors such as Jun and Fos bind to the TRE.

Transcription factors have DNA binding motifs

Many DNA-binding proteins can be grouped into classes based upon their use of structural motifs for recognition

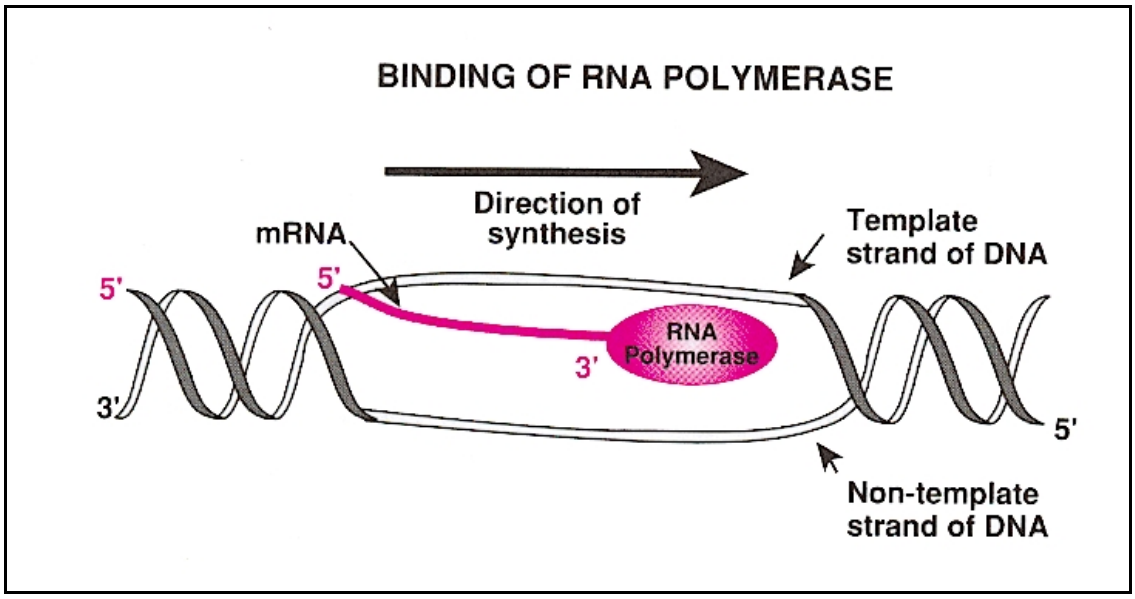
Helix-turn-helix structures - a structural motif that provides specificity of target recognition

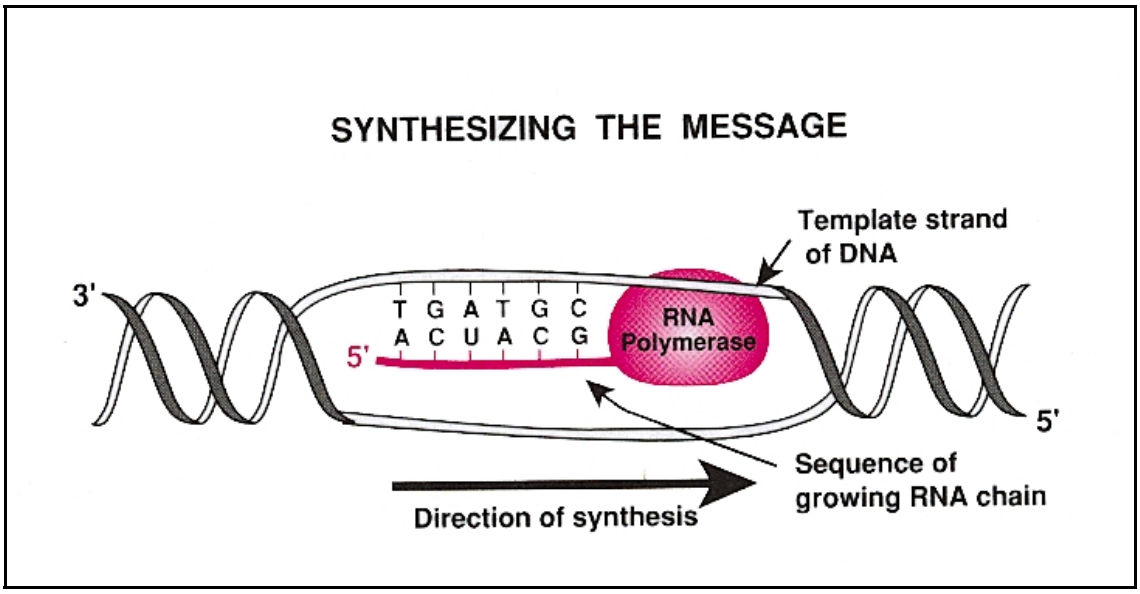
Zinc fingers - structural motif that has cysteine and histidine residues that bind zinc

Steroid receptors - regulatory proteins that include receptors for steroid hormones, retinoids, vitamin D and thyroid hormones

Leucine zippers - leucine-rich stretches of amino acids that play an important role in differentiation and development

Important in dimer formation typical of **AP1 binding proteins,** especially **homodimers** (Jun-Jun) and **heterodimers** (Jun-Fos)





Figures from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

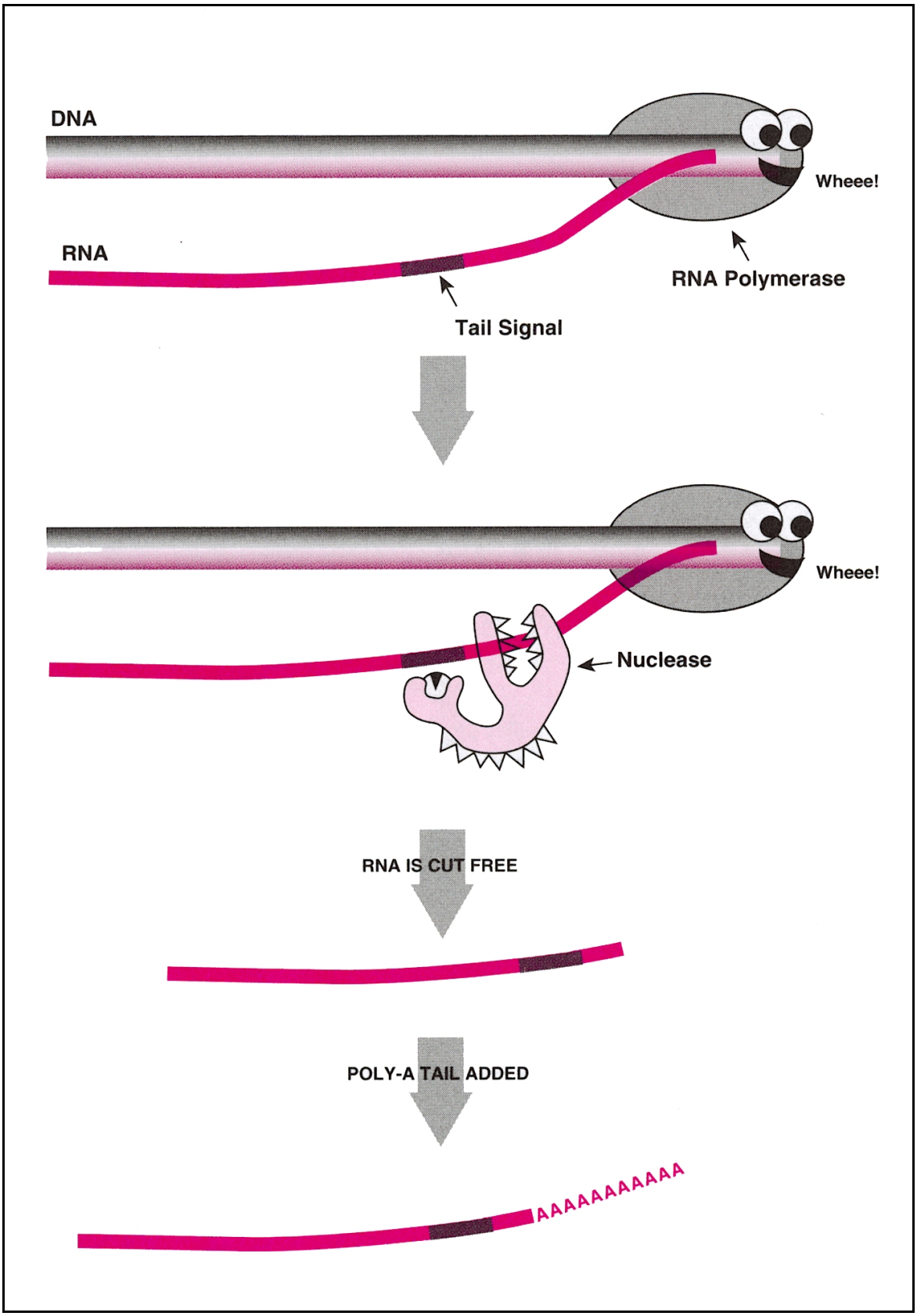


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Just as initiation of mRNA synthesis is highly regulated, the termination of mRNA synthesis is regulated by the presence of specific terminator sequences in the gene being transcribed. The signal is two inverted repeats followed by a string of A's. As a consequence of this terminator sequence,

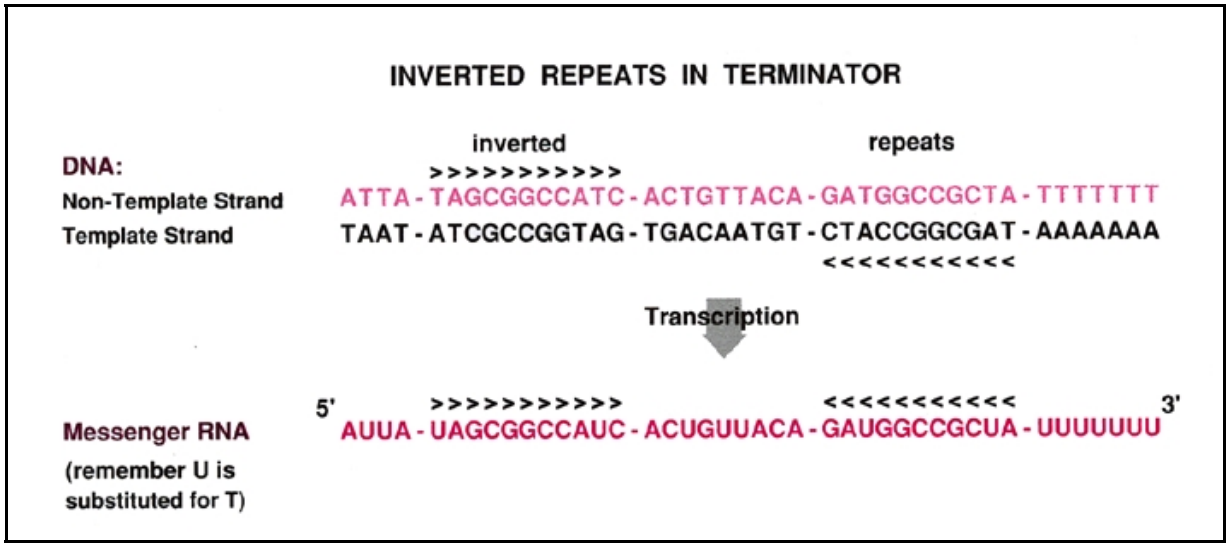


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

the mRNA contains complementary inverted repeats which form a stem and loop or hairpin structure and the string of A's in the DNA template leads to a string of U's in the mRNA.

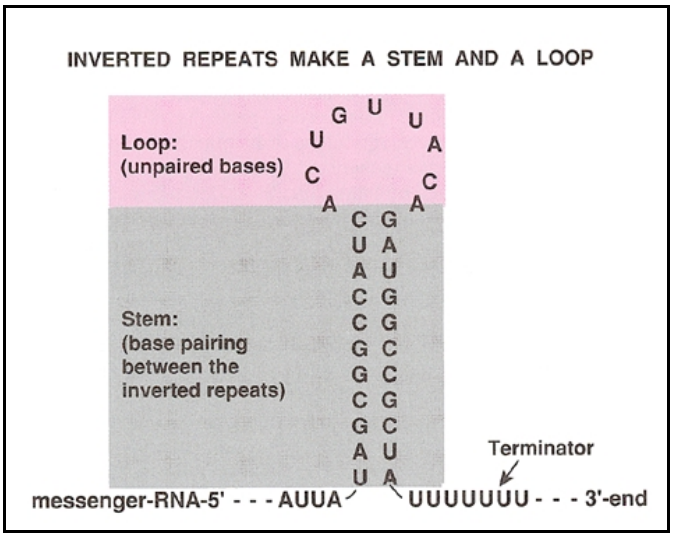


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

What basically happens is that when the RNA polymerase reaches the hairpin structure, it slows down or stops momentarily. Since the string of U's complementary to the A's forms a weak structure, the RNA and DNA fall apart, the RNA polymerase falls off, and transcription stops.

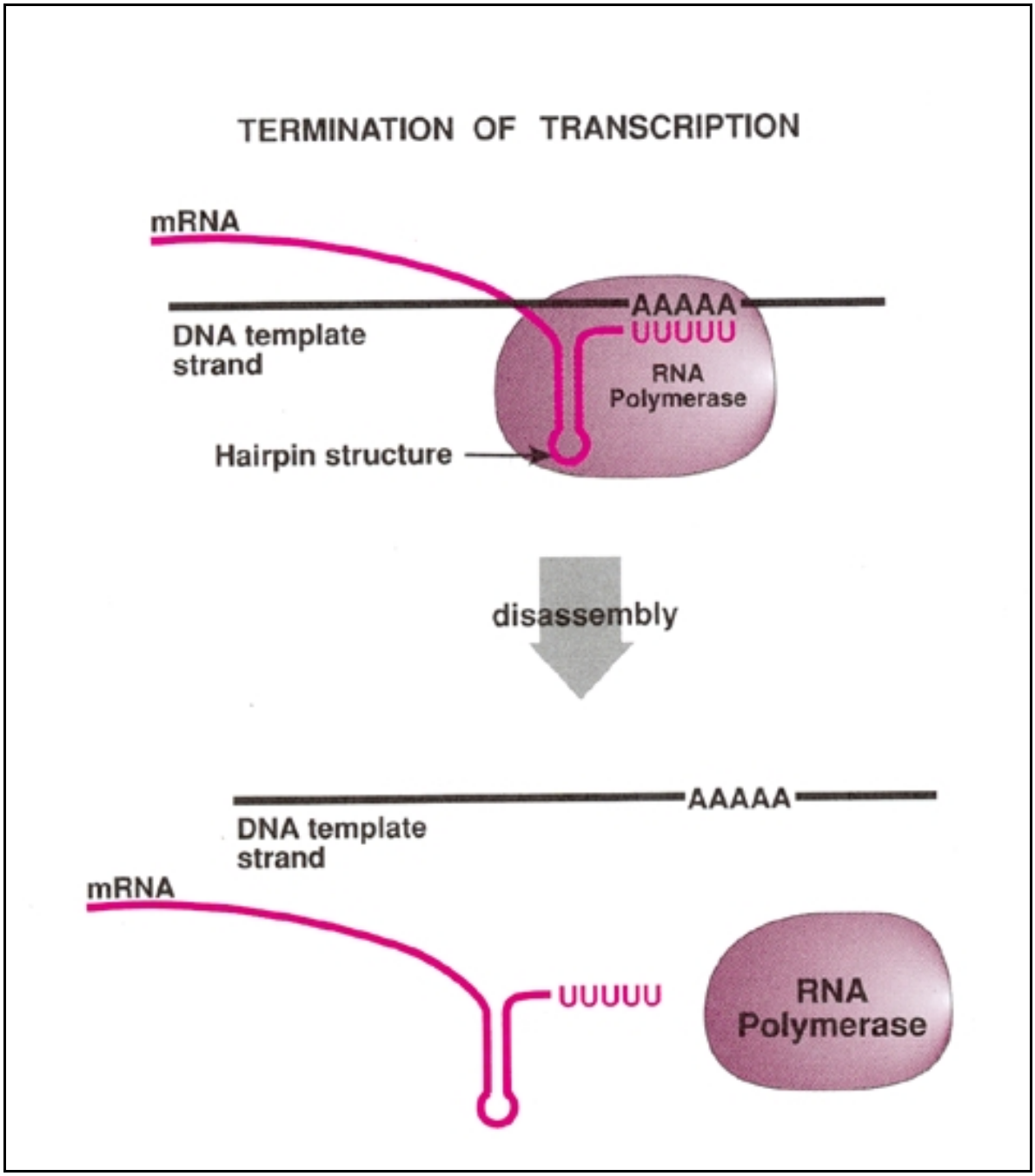


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Splicing

Splicing is the procedure in which introns which separate the coding sequences (exons) are removed in preparation for translation of the mRNA into protein. Splicing is complex and allow for creation of different proteins from the same mRNAs. Alternative splicing allows for the formation of more than one protein from a single gene.

This can occur by initiating transcription at different promoter sequences that are upstream of the gene coding sequences or by adding the polyA tail at alternative sites downstream of the gene coding sequence. Splicing can also involve exon cassette selection or trans-splicing. Trans-splicing results in mRNA derived from two separate original RNA molecules.

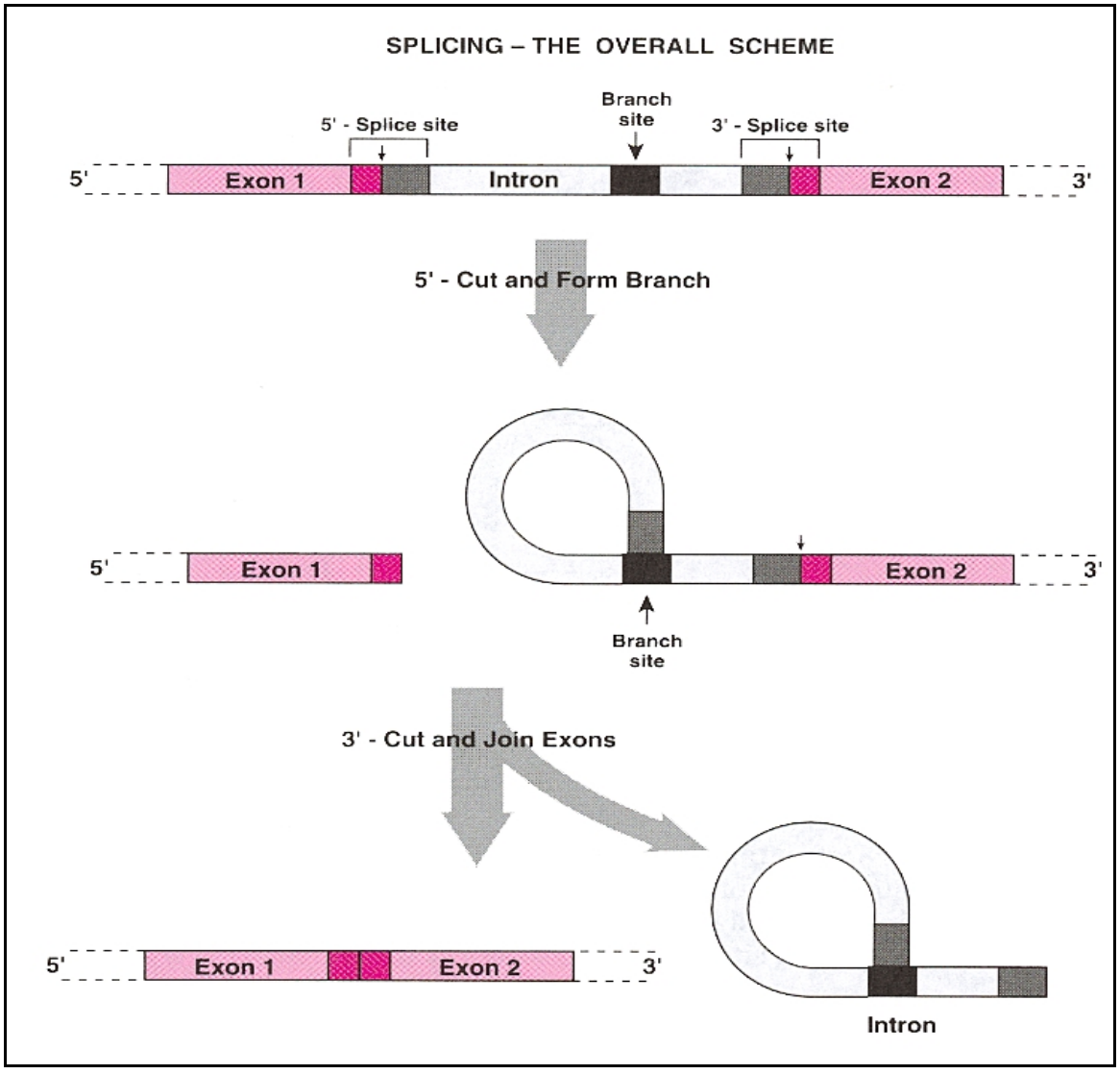


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

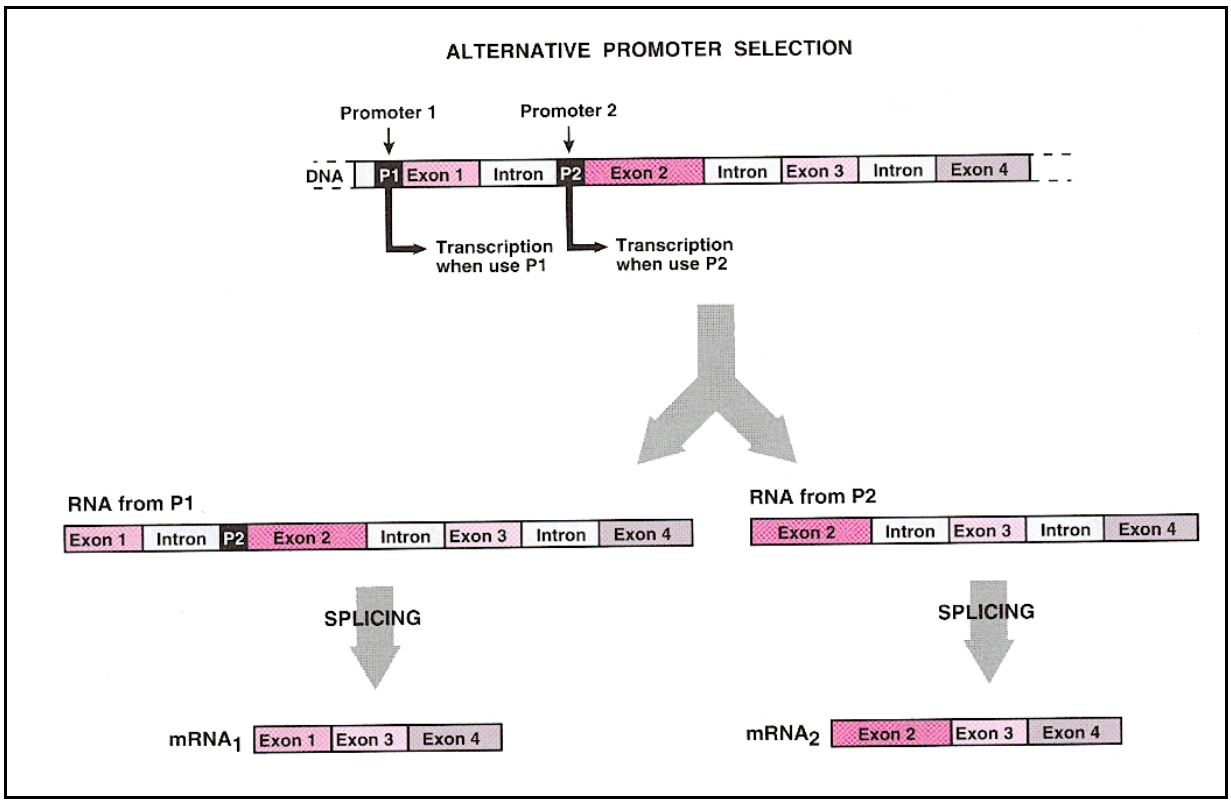


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

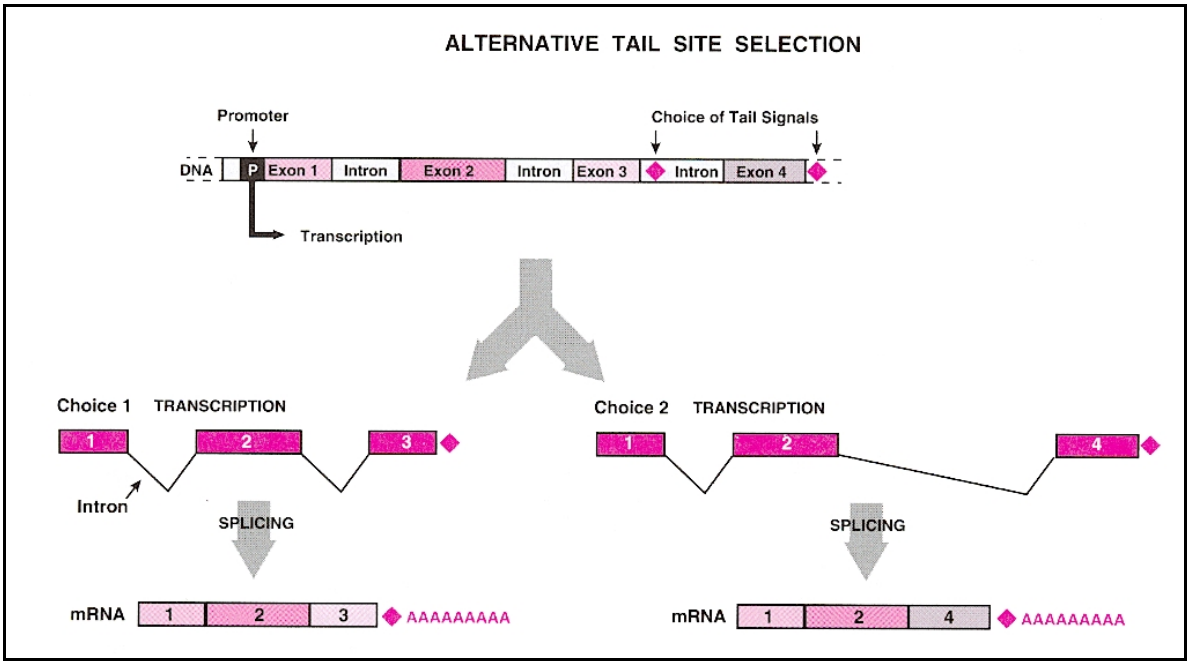


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

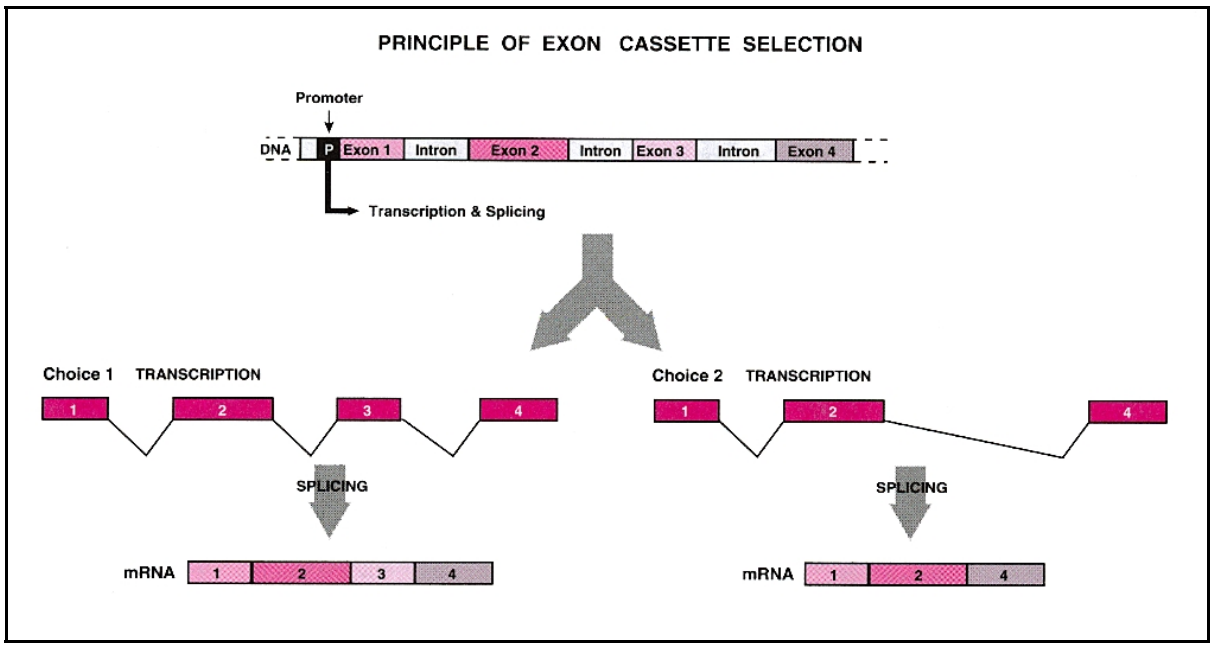


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

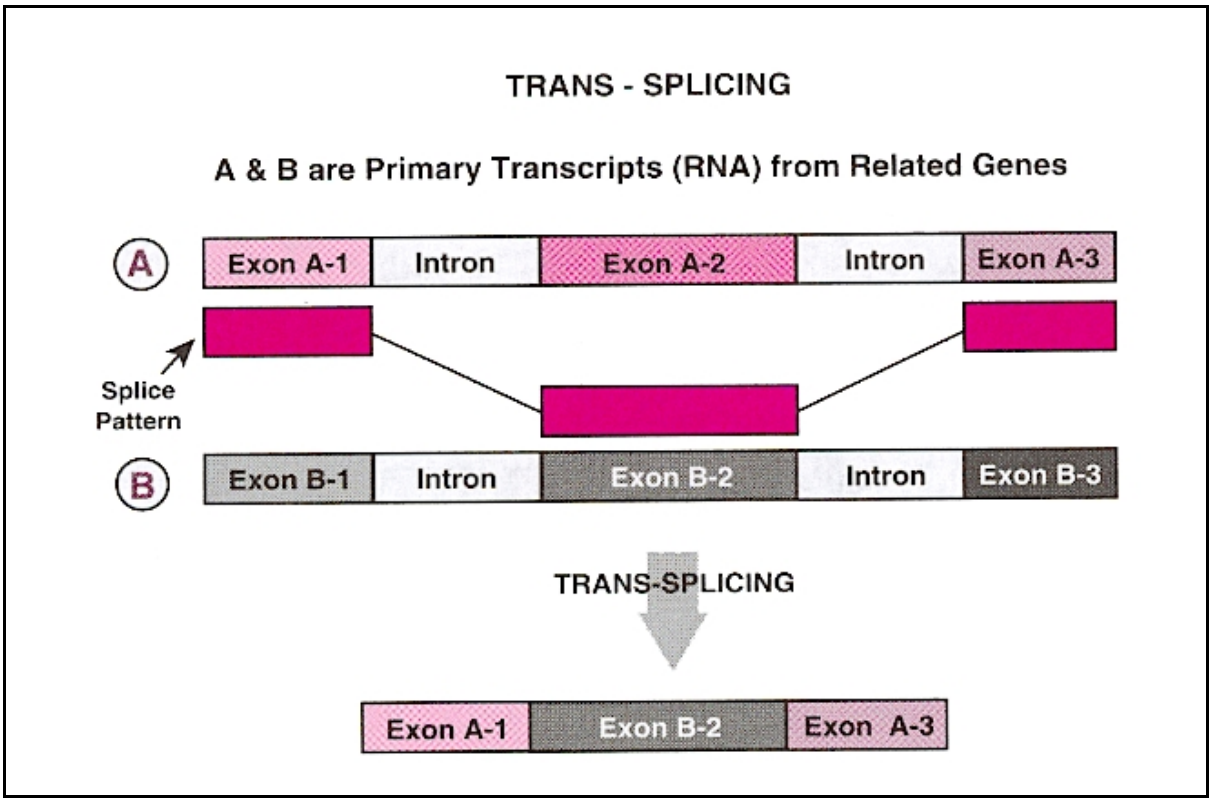


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Caps and Polyadenylation Tails

Prior to its transport from the nucleus to the cytoplasm, mRNA is capped at its 5' end and a polyA tail is added at its 3' end. The cap consists of a guanine that is added to the first nucleotide in reverse orientation via a triphosphate link. The 5' terminal guanine (G) is often methylated.

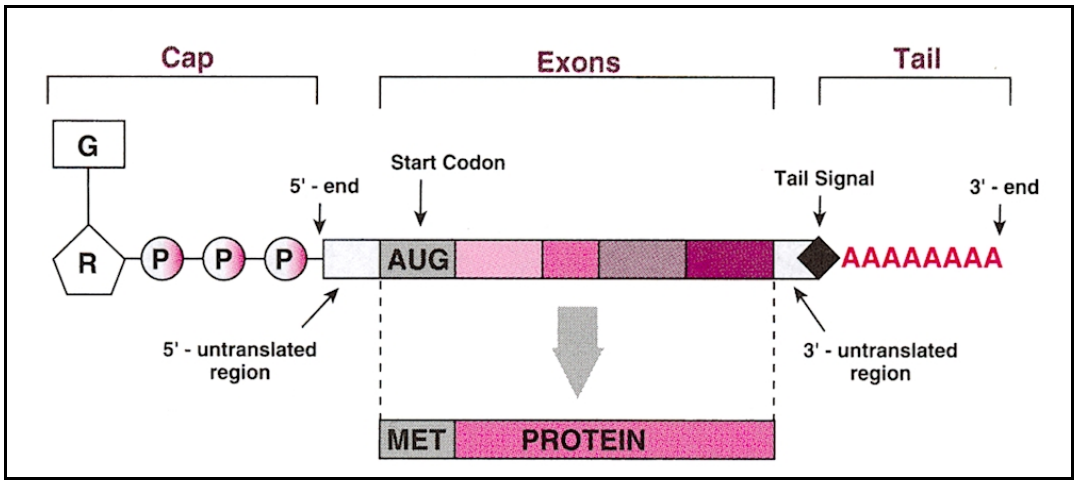


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

The cap with its complex methylated structure is believed to be important for maintaining the stability of the mRNA and is essential for translation.

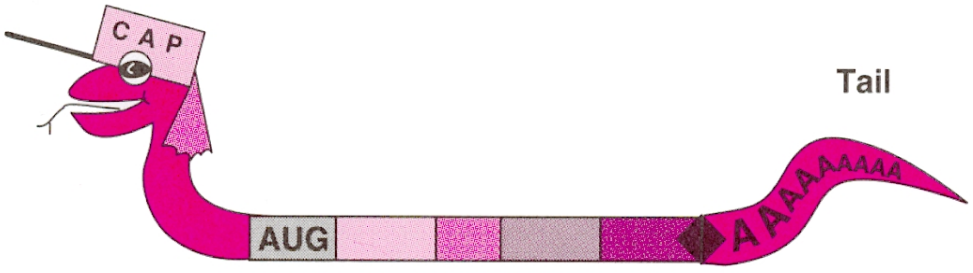


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

There is a recognition sequence at the 3' end of the mRNA molecule consisting of AAUAAA. While the RNA polymerase that is fabricating the mRNA from the DNA template continues beyond this AAUAAA recognition sequence (tail signal), a specific endonuclease recognizes this sequence and cuts the mRNA molecule at a point 10 to 30 bases downstream from the AAUAAA sequence. The enzyme poly(A) polymerase then adds a string of

100 to 300 adenine residues to form the polyA tail. The polyA tail provides the molecular biologist with a handy way to recognize the 3' end of mRNA. In terms of the biological significance of the polyA tail, little is known. It has been hypothesized that the polyA tail stabilizes the mRNA, and it has also been suggested that the polyA tail is involved in the export of mRNA from the nucleus.

Once mRNA has been capped and tailed and undergone splicing, it can then move through the nuclear pores and into the cytoplasm.

Translation of mRNA into Protein

Protein synthesis occurs in association with the ribosomes and involves several complex procedures involving rRNA and various tRNAs and accessory factors. The sequence of amino acids that constitute the protein product are derived from decoding the genetic code contained in the codons (see Genetic Code Table p. 2-7). Codons are groups of three RNA or DNA bases which encode a single amino acid. tRNAs recognize and bind to a codon on the mRNA and carry the amino acids to a ribosome. The mRNA bases are read in groups of three starting from the 5' end and always beginning with the start codon, AUG. There can be more than one start codon sequence at the 5' end of mRNA. For example, three AUG sequences are depicted below:

5' GAUAUGUAUGCGAUGCCGGAAACAUCUAAGGA 3'

Depending upon where translation starts (which start codon is used), entirely different products can result as the subsequent codons are translated in successive groups of three into specific amino acids. The three possibilities in the above example are referred to as reading frames. A sequence of mRNA that begins with a start codon and ends with a stop codon (a nonsense codon) and can thus be translated into protein is referred to as an open reading frame. Termination (stop, nonsense) codons include UAA, UAG, and UGA. All proteins start with a methionine. The codon AUG specifies methionine and is sometimes called the initiator codon. A variety of control mechanisms and the possibility of perturbations in translational control factors can undoubtedly account for cellular abnormalities. This is an area of active investigation. The stop codons signal for the termination of translation. In the absence of a stop codon (e.g., defective mRNA) the entire process may become stalled without production of a useful protein.

mRNA Degradation

Once mRNA is transcribed and translated it is subject to degradation by a regulated process. Half lives of mRNAs can vary from minutes to days. Most mRNAs need to be translated in order for there to be degradation.

Decapping enzymes, which are known to occur, could destabilize mRNA. Deadenylation of the polyA tail may be a prerequisite for mRNA degradation. Consensus sequences such as AUUUA at the 3' untranslated regions of mRNA in conjunction with 20S protein complexes have been suggested to alter cellular metabolism and precipitate degradation of mRNA.

Summary of Gene Expression

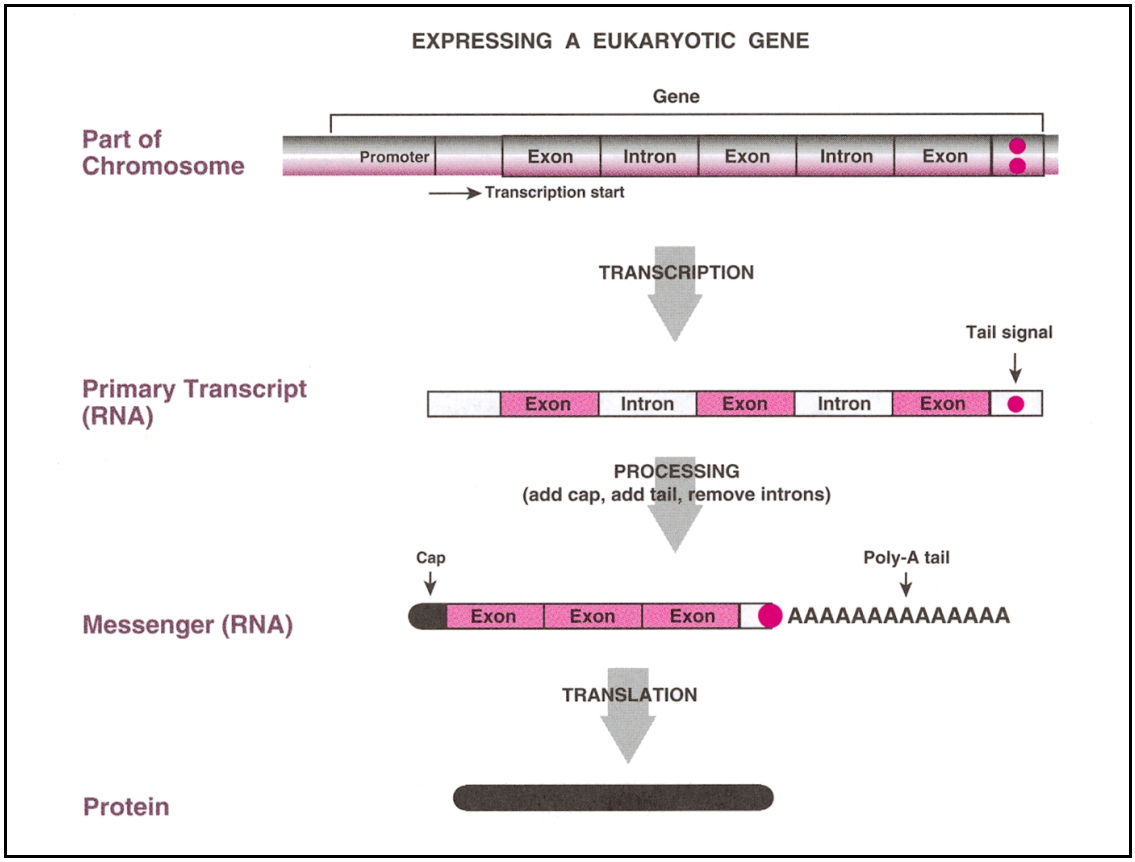


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.